SATELLITE IMAGE CLASSIFICATION USING MACHINE LEARNING

A Synopsis Submitted

To

Swami Ramanand Teerth Marathwada University, Nanded

For The

Doctor of Philosophy

Under

Computer Science and Engineering

By

Solanki Chetan Rajesh

Under the Guidance of

Dr. S. V. Bonde

Shri Guru Gobind Singhji Institute of Engineering and Technology, Vishnupuri, Nanded – 431606 (M.H.)

Table of Contents

Abstract 3

Introduction 4

Literature Survey 5

Objectives 7

Methodology 7

Impact on the society 8

Year wise plan of work and targets to be achieved 9

Expected Outcome 9

References 10

Abstract

Image classification entails the important part of digital image and has been very essential in the application of remote sensing systems, thus the demand for research to find advanced algorithms and tools to solve problems experienced in classification has shown great increase in interest over the years. Remote sensing has globally being applied with the use of current advanced satellite systems and sensors, but the need to provide analysis and decision making has been a challenge. The research will focus on devising of the algorithms for classification based on machine learning techniques. Applying machine learning techniques for the purpose of classification will be of great importance. Along with this the implementation of the newly devised algorithm(s) on GPU platforms will help in faster processing.

Introduction

Remote sensing over the years has enabled researchers, society, government and other people affected to benefit from its application on various fields of study. The possibility of collecting data and images to mapping of the earth resources, with the use of aircraft and satellites, has proven to be very beneficial to many people who needed to acquire information on earth's resources and other areas of interest.

Image classification has been very essential in the application of remote sensing systems, thus the demand for research to find advanced algorithms and tools to solve problems experienced in classification, has shown a great increase over the years. It is very important to classify images correctly because often the information derived from the analysis is utilized for vital decision making. If a wrong target is misclassified, it can lead to wrong choices that can yield disastrous results in that particular area of the problem, thus it is for this reason we are doing this type of study.

Human eye alone cannot cover large areas for analysis. Let us take a case where urban planning is needed and information obtained from the image is incorrectly classified as water instead of soil or grass that information will give wrong input for decision making. This type of problem shows that techniques that are able to classify images correctly are needed thus research in this field is very vital. Machine learning can however be able to produce techniques that can classify images better.

Classification plays an important role in mapping pixels, objects and other features found in satellite images. Machine learning can provide techniques that are able to perform the job. Slivka, (2010) defines Pattern Recognition (PR) as a branch of machine learning (ML) that focuses on the recognition of patterns and regularities in data, although in some cases, it is considered synonymous with ML. It is concerned with assigning of output values referred to as a label, to a given input value, which is referred to as an instance, by using algorithms. PR produces a platform for problem solving, in which at a later stage it s application will require a category formed into supervised classification and unsupervised classification. This means that PR tries to create a relationship between an instance and a class label. When advanced technologies and large quantities of remote sensed data are provided, it is possible to obtain better results.

Machine learning approaches are supervised learning, unsupervised learning, and semi-supervised learning approach, although the common used is the supervised and the unsupervised approaches. There are many techniques such as ANN's, SVM's, Decision Trees, Logistic Regression, K means, and many more techniques that can be used in ML. We are aware that there are other approaches, such as Deep learning and Extreme learning that are powerful tools, but for the purpose of this dissertation, we will not focus on them.

Traditional classification techniques can yield incorrect and incomplete results according to (Xu et.al. 2014 A huge effort has been taken in generating effective classification approaches and techniques that are producing better classification accuracy. ML algorithms have been researched over the years, and been identified as capable of handling RS data, with the aid of new generation satellite data. Classification methods include unsupervised and supervised approaches in RS environments.

Literature Survey

Cracknell and Hayes, (1991) discussed the different ways in which satellite systems and aircraft applications can be used and applied in fields such as land use land cover, oceanography, meteorology, climatology, natural disasters, water resources, civil engineering, and many more areas, which we will not go into detail for the purpose of this study. RS has to do with photographic instruments that can capture an image without being in contact with it. In the current study the interest is based more on image classification of satellite imagery type in RS and finding out which machine learning technique can accurately classify images.

Traditional classification techniques can yield incorrect and incomplete results according to (Xu et.al. 2014) A huge effort has been taken in generating effective classification approaches and techniques that are producing better classification accuracy. ML algorithms have been researched over the years, and been identified as capable of handling RS data, with the aid of new generation satellite data. Classification methods include unsupervised and supervised approaches in RS environments.

Camps-Valls and Bruzzone, (2009), in their book describe supervised classification as a case where a labelled set of samples, $\{(x_i, y_i)\}_i^l = 1)\}$, which is implemented on the training data by having the objective of predicting the value y corresponding to a new sample , looking at

the class membership of x. They however describe unsupervised classification method as a case that involves the type of data, (x_j), that is unknown and needs to be a way of describing how the data is organized and clustered as the major objective.

Alqurashi and Kumar, (2014) mention that in making a good decision, it is important to minimize the common sources of errors, thus the quality of thematic information and maps from remotely sensed data is vital. The problem encountered is that the accuracy results alone cannot be used as a good measure or final conclusion, in determining the correct prediction or accuracy in classification of images. The accuracy results obtained from the classification experiments, can only indicate how many times any algorithm performs accurately, but we need a way of looking at each algorithm as it performs throughout each class level of accuracy while measuring its performance. An evaluation or assessment of the performance of the supervised classifiers compared and investigated in this study, and different classifiers will be used to analyze which one will be more superior in terms of performance than the other.

The overall accuracy of the classification methods do not give satisfactory analysis in terms of evaluating comparisons of the given supervised classifiers. Lu and Weng, (2007) discuss how important it is to have a clear understanding of different sensor data, while taking into account their weaknesses and strengths in the selection of remotely sensed data, while they highlight that airborne and space borne sensor data vary in spatial, radiometric, spectral and temporal resolution. Properly selecting sensor data is considered to be the first step to a successful classification in that case. While data pre-processing is part of the process image classification, a very critical factor such as feature extraction and election is a very important aspect in selecting variables for classification of the image.

A list of potential variables for image classification which may be used is indicated by (Lu and Wu, 2007), where they mentioned spectral signatures, vegetation indices, textural, transformed images, contextual information, multi-temporal images, multisensory images and ancillary. Although various literature have indicated different approaches to performance evaluation, having more than one performance measure will yield better results.

Accuracy is affected by lots of errors found in the results of the classification experiments. Powell et al., (2004) discusses how it is important to know the sources of errors, which in other cases may include interpretation errors, poor quality of training done or errors in training samples, which all form part of factors that affect accuracy. Twala and Nkonyana, (2013) also presented a study that researched on incomplete data using supervised classification methods.

An error matrix is one of many, but mostly commonly used approaches in accuracy assessment (Lu and Weng, 2007), for which in other cases they refer to confusion matrix. In the literature review, the paper will discuss and elaborate on the different weaknesses/limitation and strengths of the different classification methods. The classification methods can be parametric or non-parametric and the study focused on the supervised ML classification techniques, in which a selection of five ML classification techniques is undertaken, in order to perform an empirical test. The experimental tests are performed on three data sets utilized against six performance measures using the Weka software. The Random Forest outperformed the other ML classification techniques.

Objectives

1. Study of machine learning algorithms/techniques.
2. The main objective of the project is to devise new algorithm(s) for the classification process by applying machine learning techniques.
3. Classification rules will be extracted from the spatial data. Then it will be applied to classify the satellite image and then the accuracy assessment of the whole process will be done to check, whether there is any change in the classification accuracy.
4. Boosting up the performance by devising the algorithms that run on GPU platforms.

Methodology

Classification of remotely sensed data is used to assign corresponding levels with respect to groups with homogeneous characteristics, with the aim of discriminating multiple objects from each other within the image.

Classification will be executed on the base of spectral or spectrally defined features, such as density, texture etc. in the feature space. Classification will be made according to the following procedures:

a) Definition of Classification Classes

b) Selection of Features

c) Sampling of Training Data

d) Estimation of Universal Statistics

e) Classification

f) Verification of Results

Impact on the society

Image classification in the RS application has a number of processes involved in acquiring the data and images obtained from aircraft and satellites. Classification of remotely sensed satellite images requires interpretation and a lot of analysis for decision-making. ML techniques play an important role in providing or giving solution to image classification problems. Biswal et al., (2013) suggests that the new generation of satellite datasets certainly demand huge computing resource and a robust classification procedure. Biswal. et al., (2013) mentions that classification algorithms are data dependent and extraction of information from such data poses various challenges. Perumal and Bhaskaran, (2010) mentions that image classification is an important part of the remote sensing data mining. He further mentions that the performance of the classifiers depends on the data. Thus, in turn he explains that it is vital to have a better understanding of data. It simply implies that the conclusion made by one algorithm in terms of being accurate can be highly challenged due to lack of enough number of experiments.

The research carried out in this dissertation can be used to benefit many institutions and professional people such as the government, social, Industry, academics and economic institutions. This work can be used to gain more analysis since it is very hard to obtain sufficient surveys like one performed in this research due to the fact that the cost, labour and experiments to perform such research are costly and time consuming. The analysis and findings of this research can provide better understanding of the current ML classification tools, and be able to establish future work for experimental aspects.

Lu and Weng, (2007) indicates that image classifications research which is based on remote sensing has for a number of years attracted the interest of the remote sensing community since most environmental and socioeconomic applications are based on the classification results.

There are other impacting issues such as unplanned settlements and increasing natural disaster that occur globally, which will need approaches such as remote sensing to better cover a large scale. This research will enable appropriate measures to be put in place, and institutions that

are involved in planning will obtain an in depth understanding of tools that are available for classification that can correctly classify an image better.

Year wise plan of work and targets to be achieved

		2017	2018				2019				2020		
SN	**Parameters**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**
1	Literature Survey												
2	Problem Definition and Research Objectives Formulation												
3	Purchase of Books and Journals												
4	Definition of Classification Classes												
5	Selection of Features												
6	Sampling of Training Data												
7	Estimation of Universal Statistics												
8	Classification												
9	Verification of Results												
10	Publications												

Expected Outcome

Satellite image classification plays a major role in extract and interpretation of valuable information from massive satellite images. The applications based on the satellite data are of great importance to the society. Accordingly efficient results of these satellite based applications is also necessary. So, this research will help in efficient classification of the satellite based data through the help of machine learning. Hence the applications built on the satellite image classification techniques will also produce the efficient results. Along with the use of GPU's these techniques will also produce faster results.

References

[1] David J. Lary, Amir H. Alavi, Amir H. Gandomi, Annette L. Walker, “Machine learning in geosciences and remote sensing”, Geoscience Frontiers, Volume 7, Issue 1, 2016, Pages 3-10, ISSN 1674-9871

[2] G. Cheng, J. Han and X. Lu, "Remote Sensing Image Scene Classification: Benchmark and State of the Art," in Proceedings of the IEEE, vol. 105, no. 10, pp. 1865-1883, Oct. 2017. doi: 10.1109/JPROC.2017.2675998

[3] Charles Toth, Grzegorz Jóźków, “Remote sensing platforms and sensors: A survey”, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 115, 2016, Pages 22-36, ISSN 0924-2716

[4] Jun Chen, Ian Dowman, Songnian Li, Zhilin Li, Marguerite Madden, Jon Mills, Nicolas Paparoditis, Franz Rottensteiner, Monika Sester, Charles Toth, John Trinder, Christian Heipke, “Information from imagery: ISPRS scientific vision and research agenda”, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 115, 2016, Pages 3-21, ISSN 0924-2716

[5] Fundamentals of Remote Sensing, Remote Sensing Tutorial by Canada Centre for Remote Sensing. Online. http://www.nrcan.gc.ca/node/9309

[6] Slivka, T., (2010). Pattern Recognition. Online: www.csun.edu/~trs05852/.

[7] Xu, L., Li, J., Brenning, A., (2014). A comparative study of different classification techniques for marine oil spill identification using RADARSAT-1 imagery. Remote Sensing of Environment, Vol. 141, pp 14-23.

[8] Valls, G.C., and Bruzzone, L., (2009). Kernel Methods for Remote Sensing Data Analysis. ISBN: 978-0-470-72211-4.

[9] Alqurashi, A.F., and Kumar, L., (2013). Investigating the Use of Remote Sensing and GIS Techniques to Detect Land Use and Land Cover Change: A review. Vol.22, No.2, pp. 193-204.

[10] Lu, D., and Weng, Q., (2007). A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing. Vol. 28, No. 5, pp. 823-870.

[11] Twala, B., Nkonyana, T.N., (2013). Extracting supervised learning classifiers from possibly incomplete remotely sensed data. 2013 1st Brics Countries on Computational Intelligence, Brics-CCI 2013, pp. 476-482.

[12] Bhaskaran, R., and Perumal, K., (2010). Supervised Classification Performance of Multispectral Images. Journal of Computing, Vol. 2, Issue 2, ISSN 2151-9617.

Solanki Chetan Rajesh

Research Scholar

Department of Computer Science and Engineering

SGGSIE&T, Nanded

Dr. S. V. Bonde

Professor

Department of Electronics and Telecommunication,

SGGSIE&T, Nanded.